Working Paper No. 154

AD699891

THE STREETWALKER'S DILEMMA: A JOB SHOP MODEL

by

STEVEN A. LIPPMAN and SHELDON M. ROSS

November, 1969

WESTERN MANAGEMENT SCIENCE INSTITUTE

University of California, Los Angeles

University of California

Los Angeles

Western Management Science Institute

Working Paper No. 154

THE STREETWALKER'S DILEMMA: A JOB SHOP MODEL

by

Steven A. Lippman[†]

and

Sheldon M. Ross[‡]

November, 1969

[†] This research was partially supported by the Western Management Science Institute under grants from the National Science Foundation, the Ford Foundation, and by an Office of Naval Research Contract, [redacted] N00014-69-A-0200-4005.

[‡] This research was partially supported by the U.S. Army Research Office--Durham, Contract DA-31-124-ARO-D-331 with the University of California.

ABSTRACT

We consider the problem of maximizing the long-run average return in a single server queueing reward system in which the customer's offer of a joint distribution of reward and service time required to earn this reward is independent of the renewal process which governs customer arrivals. After formulating the problem as a semi-Markov decision process, we characterize the form of an optimal policy. When the renewal process is Poisson, the characterization is easily stated: accept a customer if and only if the ratio of his expected reward to his expected service time is larger than g, the long-run average return. When the arrival process is Poisson, g is easily found. Next, batch arrivals are permitted, and further results are obtained.

1. Introduction

We consider the problem of maximizing the long-run average return in a single server queueing reward system in which the customer's offer of a joint distribution of reward and service time required to earn this reward is independent of the renewal process which governs customer arrivals. In describing the model, we find it enlightening to introduce the necessary notation and terminology in the context of a problem which we refer to as "the streetwalker's dilemma."

Consider a streetwalker working in a large city, and suppose that her customers arrive according to a renewal process having interarrival distribution F with $F(0) = 0$. Each arriving customer makes an offer which she must either accept or reject, and all customers who arrive while she is busy or whose offer she has rejected are assumed lost. Thus pre-emption and backlogging are not permitted. If she accepts a customer (i.e., an offer) of type x, $-\infty < x < \infty$, then the probability that she will receive no more than s dollars and that the service time required to earn this reward will not exceed t is given by the joint distribution $G_x(s,t)$. Furthermore, the distribution function H of the type of offer she receives is independent of the renewal process and of her past decisions, and hence successive offers are independent and identically distributed. The streetwalker's dilemma, then, is to decide which customers to accept and which customers to reject so as to maximize her long-run average return.

The model can be viewed as one for determining whether or not a factory or job shop should accept potential jobs. Several other interesting

examples of this model are given by Miller [2, pp. 67-70]. The fundamental difference between our model and Miller's [2] is that his is restricted to (i) exponential service time which is assumed independent of the customer type, (ii) Poisson arrivals, and (iii) a finite number of customer types. On the other hand, Miller has the added generality of allowing many servers.

In section 2 we formulate the problem as a semi-Markov decision process and introduce the necessary notation. Employing recent results due to Ross [4], we determine the structure of an optimal policy in section 3. Next, we specialize to the case of Poisson arrivals and prove a monotonic property which enables us to easily calculate, in practice, the optimal policy. Finally, we allow batch arrivals, and again determine the structure of an optimal policy.

2. Notation and Definitions

In characterizing the structure of an optimal policy, it behooves us to formulate our model as a semi-Markov decision process.

Definition. A semi-Markov decision process is a process with state space S and action space A. Whenever a transition to some state occurs, an action is chosen. If the state is x and action a is chosen, then

(i) the next state of the system is given by the distribution function $P_{x,a}(\cdot)$,

(ii) conditional on the event that the next state is y, the time until the transition from x to y occurs is a random variable with distribution function $F_{x,y,a}(\cdot)$,

and

(iii) there is a reward earned at the time the action is taken, and it is a random variable depending on x and a.

When the transition times are identically one, that is, when $F_{x,y,a}(t)$ equals 0 for $t < 1$ and 1 for $t \geq 1$, then we have the more familiar Markov decision process.

We shall say that an offer is made only when a customer arrives and finds the streetwalker free. Then in our model, the process is said to be in state x at time t if the last offer made at or prior to time t was an offer of type x so $S = (-\infty,\infty)$. We say that a transition to state y occurs at time t if an offer of type y is made at time t. In each state there are two possible actions: accept (action 1) and reject (action 2). Thus, $A = \{1,2\}$. The reward earned for rejecting an offer is zero while the reward earned for accepting an offer of type x has distribution function $R_x(\cdot)$ given by

$$R_x(s) = \int_0^\infty dG_x(s,t) \; .$$

The transition function $P_{x,a}$ is independent of x and a and equals H, whereas the distribution $F_{x,y,a}$ of time until the next transition occurs equals F if $a = 2$. We give $F_{x,y,a}$ later for the case $a = 1$.

A <u>policy</u> π is any (possibly randomized) rule which for each $t \geq 0$ specifies which action to take at time t given the current state and the past decisions and history of the process. Of particular interest are (nonrandomized) stationary policies which, independently of the time t and the past decisions and history of the process, simply specify which action to take from each state. In our model, a <u>stationary policy</u>

separates the types of offers into two categories: those we always accept and those we always reject.

For each policy π and each state x, we define

(1) $$\varphi_\pi^1(x) = \liminf_{t \to \infty} E_\pi\left[\frac{Z(t)}{t} \mid X_1 = x\right]$$

and

(2) $$\varphi_\pi^2(x) = \liminf_{n \to \infty} \frac{E_\pi\left[\sum_{j=1}^{n} Z_j \mid X_1 = x\right]}{E_\pi\left[\sum_{j=1}^{n} \tau_j \mid X_1 = x\right]}$$

where $Z(t)$, Z_j, and τ_j denote, respectively, the total rewards received by time t, the reward received during the j^{th} transition interval, and the length of the j^{th} transition interval; X_1 is the initial state.

The criterion given in (1) is the usual definition of the long-run average return. The criterion given in (2) was first suggested by Ross [4] and is the limit of the ratio of the expected reward earned during the first n transitions to the expected time for the first n transitions. Even though (1) is slightly more appealing than (2), we shall adopt (2) as our definition of the long-run average return as it is more amenable to analysis. We shall show, however, that the two criteria are equivalent for stationary policies.

The problem, then, is to find a policy π^*, termed <u>optimal</u>, such that

$$\varphi_{\pi^*}^2(x) = \sup_\pi \varphi_\pi^2(x), \quad \text{for each } x \in S .$$

Finally, let $\bar{R}(x,1)$ and $\bar{R}(x,2)$ denote the expected reward received during a transition interval which begins with her acceptance or rejection of an offer of type x, respectively. Also, denote by $\bar{\tau}(x,1)$ and $\bar{\tau}(x,2)$ the expected length of a transition interval which begins with her acceptance or rejection of an offer of type x, respectively. Then

$$\bar{R}(x,1) = \int_0^\infty \int_{-\infty}^\infty s\,dG_x(s,t) ,$$

$$\bar{R}(x,2) = 0 ,$$

$$\bar{\tau}(x,1) = \int_0^\infty \int_{-\infty}^\infty (t + EY_t)\, dG_x(s,t) ,$$

and

$$\bar{\tau}(x,2) = \int_0^\infty y\,dF(y) \equiv \mu ,$$

where EY_t is the expected amount of time that she must wait (remain idle) until she receives another offer given that she spent an amount of time t with her previous customer. (Y_t is just the excess life at time t of the renewal process [3, p. 173].) Note too that $\tau(x,1) > \mu$ for all x.

3. Optimal Policies

Of considerable importance is the fact [4] that we can assume without loss of generality that whenever action a is taken in state x, then the length of each transition interval is identically $\bar{\tau}(x,a)$ and the reward earned is identically $\bar{R}(x,a)$. Using this fact, it follows

that $V_\alpha(x,n)$, the maximal expected α-discounted reward earned during the first n transitions when $X_1 = x$, is given by $(V_\alpha(x,0) \equiv 0)$

$$(3) \qquad V_\alpha(x,n) = \max\left\{\overline{R}(x,1) + e^{-\alpha\overline{\tau}(x,1)}\int_{-\infty}^{\infty} V_\alpha(y,n-1)dH(y)\ ;\ e^{-\alpha\mu}\int_{-\infty}^{\infty} V_\alpha(y,n-1)dH(y)\right\} .$$

Note that $V_\alpha(x,n)$ is increasing in n for each x.

Throughout the remainder of the paper, we shall assume that the following condition holds:

Condition 1: There is an $M < \infty$ such that $|\overline{R}(x,1)| \le M$ for all $x \in S$.

Lemma 1. The limit function $V_\alpha(x) \equiv \lim_{n\to\infty} V_\alpha(x,n)$ exists, is bounded in x, and satisfies the functional equation

$$(4) \qquad V_\alpha(x) = \max\left\{\overline{R}(x,1) + e^{-\alpha\overline{\tau}(x,1)}\int_{-\infty}^{\infty} V_\alpha(y)dH(y);\quad e^{-\alpha\mu}\int_{-\infty}^{\infty} V_\alpha(y)dH(y)\right\} .$$

Proof. Assume that $V_\alpha(x,n-1) \le M(1-e^{-\alpha n\mu})/(1-e^{-\alpha\mu})$. Then since $V_\alpha(x,n-1) \ge 0$, $\overline{\tau}(x,1) \ge \mu$ for each x, and $R(x,1) \le M$, we have

$$\begin{aligned} V_\alpha(x,n) &\le \max\,\{\overline{R}(x,1),0\} + e^{-\alpha\mu}\int_{-\infty}^{\infty} V_\alpha(y,n-1)dH(y) \\ &\le M + e^{-\alpha\mu}\sup_y V_\alpha(y,n-1) \\ &\le M + e^{-\alpha\mu}M(1-e^{-\alpha n\mu})/(1-e^{-\alpha\mu}) \\ &= M(1-e^{-\alpha(n+1)\mu})/(1-e^{-\alpha\mu}) . \end{aligned}$$

Thus, it follows that $V_\alpha(x,n)$ is uniformly bounded in x and n. Therefore, we can conclude that the limit exists and is bounded in x since $V_\alpha(x,n)$ is increasing in n for each x. The desired result now follows by applying the Lebesque dominated convergence theorem to (3).

Q.E.D.

Lemma 2. For each pair x,z in S and all $\alpha > 0$, $|V_\alpha(x)-V_\alpha(z)| \leq M$.

Proof. Fix $x,z \in S$ and $\alpha > 0$, recall that $\overline{\tau}(x,1) \geq \mu$, and note that $\int_{-\infty}^{\infty} V_\alpha(y)dH(y) \geq 0$. By Lemma 1, $V_\alpha(z) \geq e^{-\alpha\mu}\int_{-\infty}^{\infty} V_\alpha(y)dH(y)$ so again by Lemma 1 we have either

$$
\begin{aligned}
V_\alpha(x) &= \overline{R}(x,1) + e^{-\alpha\overline{\tau}(x,1)}\int_{-\infty}^{\infty} V_\alpha(y)dH(y) \\
&\leq M + e^{-\alpha\overline{\tau}(x,1)}\int_{-\infty}^{\infty} V_\alpha(y)dH(y) \\
&\leq M + V_\alpha(z) \ ,
\end{aligned}
$$

or

$$V_\alpha(x) = e^{-\alpha\mu}\int V_\alpha(y)dH(y) \leq V_\alpha(z) \ .$$

In either case, we have $V_\alpha(x) - V_\alpha(z) \leq M$ so the desired result is obtained by reversing the roles of x and z.

Q.E.D.

Theorem 3. It is optimal to accept an offer of type x if and only if

$$\frac{\overline{R}(x,1)}{\overline{\tau}(x,1) - \mu} \geq g \ ,$$

where g is the optimal long-run average reward, i.e., $g = \sup_\pi \phi_\pi^2(x)$ for all x.

Proof. It follows from Lemma 2 and Theorem 3 of reference 4 that there is a bounded function h and a constant g such that

$$(5)\quad h(x) = \max\left\{\bar{R}(x,1) + \int_{-\infty}^{\infty} h(y)dH(y) - g\bar{\tau}(x,1);\ \int_{-\infty}^{\infty} h(y)dH(y) - g\mu\right\},\ \text{for all } x .$$

Finally, Theorem 2 of reference 4 states that if there is a bounded function h and a constant g which satisfies (5), then there is a stationary policy π^* such that

$$g = \varphi^2_{\pi^*}(x) = \max_{\pi} \varphi^2_{\pi}(x) \quad \text{for all } x ;$$

and for each x, π^* prescribes an action which maximizes the right side of (5).

Q.E.D.

It can be shown [4, p. 5] that if the expected length of a transition interval is finite for a stationary policy π, then $\varphi^1_{\pi} \equiv \varphi^2_{\pi}$. Moreover, in view of condition 1, it is easy to show by a simple renewal reward argument (see [1]) that if the expected length of a transition interval is infinite for a stationary policy π, then $\varphi^1_{\pi} \equiv \varphi^2_{\pi} \equiv 0$. This establishes Theorem 4.

Theorem 4. For each stationary policy π, $\varphi^1_{\pi} \equiv \varphi^2_{\pi}$. Hence, a best stationary policy in the sense of (1) is given by Theorem 3.

Poisson Arrivals

Of particular interest, is the special case wherein the renewal process of arrivals is a Poisson process with rate λ. Here, $\mu = 1/\lambda$, and it follows from the memoryless property of Poisson processes that

$EY_t = 1/\lambda$ for each t. Hence,

$$\bar{\tau}(x,1) = t_x + 1/\lambda ,$$

where

$$t_x = \int_0^\infty \int_{-\infty}^\infty t\,dG_x(s,t)$$

is the mean time that the streetwalker spends with a customer of type x. Theorem 3 now simplifies and takes on a more intuitive form: it is optimal to accept an offer of type x if and only if $\bar{R}(x,1)/t_x \geq g$, that is, if and only if the ratio of the mean reward to the mean service time is at least as large as the long-run average reward.

Although we have determined the structure of an optimal policy, it remains to determine g. We now establish a monotonic property which, in practice, enables us to easily calculate g.

<u>Theorem 5</u>. Suppose the arrival (renewal) process is Poisson, and let $g(c)$ be the long-run average reward when an offer of type x is accepted if and only if $\bar{R}(x,1)/t_x \geq c$. Then $g(\cdot)$ is unimodal.

<u>Proof</u>. Let $R = \{x : \bar{R}(x,1)/t_x \geq c\}$, $R' = \{x : \bar{R}(x,1)/t_x \geq c'\}$, and $p = \int_R dH(x)$. Using abbreviated notation, we have (see [11])

$$(6) \qquad g(c) = \frac{(1-p)\cdot 0 + p\int_R \frac{\bar{R}}{p}dH}{(1-p)\mu + p(\mu + \int_R \frac{t}{p}dH)} = \frac{\int_R \bar{R}}{\mu + \int_R t} .$$

Fix $c' > c$. Upon considerable rearranging of terms, we obtain[1/]

[1/] We assume that $\int_{R-R'} dH > 0$ otherwise it is obvious that $g(c) = g(c')$.

$$\text{sign}\,[g(c') - g(c)] = \text{sign}\left[g(c') - \frac{\int_{R\sim R'} \bar{R}}{\int_{R\sim R'} t}\right] . \tag{7}$$

Also, the definition of R and R' yields

$$c \le \frac{\int_{R\sim R'} \bar{R}}{\int_{R\sim R'} t} < c' . \tag{8}$$

If $c \ge g$, then (7) and (8) yield

$$g(c') - \frac{\int_{R\sim R'} \bar{R}}{\int_{R\sim R'} t} \le g(c') - c \le g - c \le 0 ,$$

so $g(\cdot)$ is nonincreasing on $[g,\infty)$.

To show g is nondecreasing on $(-\infty,g]$, it suffices by (7) and (8) to show that $g(c) \ge c$ for all $c \le g$. By definition of $g(c)$,

$$\begin{aligned}
[\mu + \int_R t][g(c) - c] &= \int_R (\bar{R} - ct) - c\mu \\
&= \int_{R\sim R''} (\bar{R} - ct) + \int_{R''} \bar{R} - c(\mu + \int_{R''} t) \\
&= \int_{R\sim R''} (\bar{R} - ct) + \int_{R''} (1 - \frac{c}{g})\, \bar{R} \ge 0 ,
\end{aligned}$$

where $R'' = \{x : \bar{R}(x,1)/t_x \ge g\}$. The last equality follows from (6).

Q.E.D.

Batch Arrivals

Suppose that customers arrive in batches. In particular, suppose that (i) each arrival consists of n customers with probability p_n and $\sum_{n=1}^{N} p_n = 1$, (ii) the batch size and the offers are independent of the renewal process, and (iii) the conditional distribution of offers given that there are n customers in the batch is given by H_n, so

$$H_n(x_1,\ldots,x_n) = P\{i^{th} \text{ offer is } \le x_i,\ i = 1,2,\ldots,n \mid \text{batch size is } n\} .$$

As before, the streetwalker can accept at most one customer from each batch, and all rejected offers are lost. Transitions are defined as before, and we say that the system is in state $(x_1,\ldots,x_n) \in \mathcal{X}$ if the last offer made was the batch $x_1,\ldots,x_n$ of offers.

Recall that in the special case of Poisson arrivals it is optimal to accept an offer of type x if and only if $\rho_x \ge g$ where $\rho_x \equiv \bar{R}(x,1)/t_x$. Consequently, this leads one to the conjecture that with batch and Poisson arrivals, the streetwalker accepts that offer--if any-- x_{i*} for which $\rho_{x_{i*}} = \max\{\rho_{x_i} : i = 1,2,\ldots,n\}$. This conjecture is, however, false, for it turns out that the relevant quantity is $\bar{R}(x,1) - t_x g$ rather than ρ_x. Thus, offers cannot be ranked according to ρ_x even though ρ_x provides us with a simple acceptance-rejection criterion.

Theorem 6. When batch arrivals are permitted, it is optimal to reject all offers from the batch $x_1,\ldots,x_n$ if and only if

$$\frac{\bar{R}(x_i,1)}{\tau(x_i,1) - \mu} < g \quad \text{for } i = 1,2,\ldots,n .$$

If an offer from the batch $x_1,\dots,x_n$ is accepted, then it is optimal to accept that offer with the largest value of

$$\overline{R}(x_i,1) - g\overline{\tau}(x_i,1) .$$

<u>Proof</u>. The arguments used in establishing Lemmas 1 and 2 suffice to establish their analogues for the case of batch arrivals. Hence, there is a bounded function h and a constant g such that for all $(x_1,\dots,x_n) \in \mathfrak{X}$ we have

$$(9)\quad h(x_1,\dots,x_n) = \max\Big\{ \max_{i=1,2,\dots,n} \Big[\overline{R}(x_i,1) + \sum_{n=1}^{N} p_n \int h(y_1,\dots,y_n)\,dH_n(y_1,\dots,y_n) - g\overline{\tau}(x_i,1)\Big] ;\ \sum_{n=1}^{N} p_n \int h(y_1,\dots,y_n)\,dH_n(y_1,\dots,y_n) - g\mu \Big\} .$$

The desired results now follow as shown in the proof of Theorem 3.

Q.E.D.

REFERENCES

[1] Johns, M. and R. Miller, "Average Renewal Loss Rate," <u>Annals of Mathematical Statistics</u>, 34, 396-401, 1963.

[2] Miller, Bruce, "Finite State Continuous Time Markov Decision Processes with Applications to a Class of Optimization Problems in Queueing Theory," Technical Report No. 1[illegible], March 10, 1967, Operations Research Department, Stanford University, Stanford, California.

[3] Prabhu, N. U., <u>Stochastic Processes</u>, Macmillan, New York, 1965.

[4] Ross, Sheldon M., "Average Cost Semi-Markov Decision Processes," Operations Research Center Report ORC 69-27, University of California, Berkeley, September 1969.

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Western Management Science Institute

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
THE STREETWALKER'S DILEMMA: A JOB SHOP MODEL

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)
Steven A. Lippman and Sheldon M. Ross

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
November, 1969	15	4

8a. CONTRACT OR GRANT NO.

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Working Paper No. 154

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. DISTRIBUTION STATEMENT
Distribution is Unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Western Management Science Institute
University of California
Los Angeles, California 90024

13. ABSTRACT

We consider the problem of maximizing the long-run average return in a single server queueing reward system in which the customer's offer of a joint distribution of reward and service time required to earn this reward is independent of the renewal process which governs customer arrivals. After formulating the problem as a semi-Markov decision process, we characterize the form of an optimal policy. When the renewal process is Poisson, the characterization is easily stated: accept a customer if and only if the ratio of his expected reward to his expected service time is larger than g, the long-run average return. When the arrival process is Poisson, g is easily found. Next, batch arrivals are permitted, and further results are obtained.

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Job shop models Queueing reward system Semi-Markov decision processes						